SARS-CoV-2 Entry Factors are Highly Expressed in Nasal Epithelial Cells Together with Innate Immune Genes

Waradon Sungnak[1,†], Ni Huang[2], Christophe Bécavin[3], Marijn Berg[4, 5], Rachel Queen[6], Monika Litvinukova[7,8], Carlos Talavera-Lopez[9], Henrike Maatz[10], Daniel Reichart[11], Fotios Sampaziotis[12,13,14], Kaylee B. Worlock[15], Masahiro Yoshida[16], Josephine L. Barnes[17], and HCA Lung Biological Network[*,†]

[1]Wellcome Sanger Institute, Cambridge, UK

[2]Wellcome Sanger Institute, Cambridge, UK

[3]Université Côte d'Azur, CNRS, IPMC, Sophia-Antipolis, France

[4]Department of Pathology and Medical Biology, University Medical Centre Groningen, University of Groningen, Groningen, Netherlands

[5]Groningen Research Institute for Asthma and COPD, University Medical Centre Groningen, University of Groningen, Groningen, Netherlands

[6]Bioinformatics Core Facility, Newcastle University Biosciences Institute, Faculty of Medical Sciences, Newcastle University, Newcastle-upon-Tyne, UK

[7]Wellcome Sanger Institute, Cambridge, UK

[8]Cardiovascular and Metabolic Sciences, Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany

[9]Wellcome Sanger Institute, Cambridge, UK

[10]Cardiovascular and Metabolic Sciences, Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany

[11]Department of Genetics, Harvard Medical School, Boston, MA 02115, United States

[12]Wellcome and MRC Cambridge Stem Cell Institute, University of Cambridge, Jeffrey Cheah Biomedical Centre Biomedical Campus, Puddicombe Way, Cambridge CB2 0AW, UK

[13]Department of Medicine, Addenbrookes Hospital, Box 157, Hills Road, Cambridge CB2 0QQ, UK

[14]Cambridge Liver Unit, Cambridge University Hospitals, Cambridge Biomedical Campus, Hills Road, Cambridge CB2 0QQ, UK

[15]UCL Respiratory, Division of Medicine, University College London, WC1E 6JF, London, UK

[16]UCL Respiratory, Division of Medicine, University College London, WC1E 6JF, London, UK

[17]UCL Respiratory, Division of Medicine, University College London, WC1E 6JF, London, UK

[†]Authors for correspondence (ws4@sanger.ac.uk; lung@humancellatlas.org)

Abstract

We investigated SARS-CoV-2 tropism by surveying expression of viral entry-associated genes in single-cell RNA-seq data from multiple tissues from healthy human donors. We co-detected these transcripts in specific respiratory, corneal, and intestinal epithelial cells, potentially explaining the high efficiency of SARS-CoV-2 transmission. These genes are co-expressed in nasal epithelial cells with genes involved in innate immunity, highlighting the cells' potential role in initial viral infection, spread and clearance. The study offers a useful resource for further lines of inquiry with valuable clinical samples from COVID-19 patients, and we provide our data in a comprehensive, open, and user-friendly fashion at covid19cellatlas.org.

Main

The coronavirus disease 2019 (COVID-19) is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)[1]. Detection of the virus was first reported in Wuhan[2], China and has since spread worldwide, emerging as a global pandemic[3].

In symptomatic patients, nasal swabs have yielded higher viral loads than throat swabs[4]. The same distribution was observed in an asymptomatic patient[4], implicating the nasal epithelium as a portal for initial infection and transmission. Cellular entry of coronaviruses depends on the binding of the spike (S) protein to a specific cellular receptor and subsequent S protein priming by cellular proteases. Similar to SARS-CoV[5,6], SARS-CoV-2 employs ACE2 as a receptor for cellular entry. The binding affinity of the S protein and ACE2 was found to be a major determinant of SARS-CoV replication rates and disease severity[4,7]. Viral entry also depends on TMPRSS2 protease activity, and cathepsin B/L activity may be able to substitute for TMPRSS2[7].

ACE2 and TMPRSS2 have been detected in both nasal and bronchial epithelium by immunohistochemistry[8]. Gene expression of *ACE2* and *TMPRSS2* has been reported to occur largely in alveolar epithelial type II (AT-2) cells[9-11], which are central to SARS-CoV pathogenesis, while a different study reported the absence of *ACE2* in the upper airway[12]. To clarify the expression patterns of *ACE2* and *TMPRSS2*, we analyzed their expression and the expression of other genes potentially associated with SARS-CoV-2 pathogenesis at cellular resolution, using scRNA-seq datasets from healthy donors generated by the Human Cell Atlas consortium and other resources to inform and prioritize the use of precious, limited clinical material that is becoming available from COVID-19 patients.

We investigated the gene expression of *ACE2* in multiple scRNA-seq datasets from different tissues, including those of the respiratory tree, cornea, retina, esophagus, ileum, colon, heart, skeletal muscle, spleen, liver, placenta/decidua, kidney, testis, pancreas, prostate gland, brain, skin, and fetal tissues. We note that studies may lack specific cell types due to their sparsity, the challenges associated with isolation, or analysis methodology. Moreover, expression may be under-detected due to technical dropout effects. Thus, while positive (presence) results are highly reliable, absence should be interpreted with care.

ACE2 expression was generally low in all analyzed datasets. Consistent with independent studies[10,11], *ACE2* was expressed in cells from multiple tissues, including airways, cornea, esophagus, ileum, colon, liver, gallbladder, heart, kidney, and testis (**Fig. 1a**; first column). *TMPRSS2* was highly expressed with a broader distribution (**Fig. 1a**; second column), suggesting that *ACE2*, rather than *TMPRSS2*, may be a limiting factor for viral entry at the initial infection stage. Cells from the respiratory tree, cornea, esophagus, ileum, colon, gallbladder, and common bile duct expressed both genes in the same cell (**Fig. 1a**; third column). We also assessed *ACE2* and *TMPRSS2* expression in developmental datasets from fetal tissues, including liver, thymus, skin, bone marrow, yolk sac, and lung, and found little to no expression of *ACE2* in all but fetal liver and thymus (**Fig. 1a**) where there was no co-expression with *TMPRSS2* (data not shown) except for a cluster of medullary thymic epithelial cells (**Fig. 1a**). *ACE2* expression is noticeable in certain cell types in placenta/decidua without *TMPRSS2* (**Fig. 1a**). Additional fetal data across relevant tissues and stages are needed to determine the generality of these findings.

To further characterize specific epithelial cell types expressing *ACE2*, we evaluated the *ACE2* expression within the lung and airway epithelium. We found that, despite a low level of expression overall, *ACE2* was expressed in multiple epithelial cell types across the airway, as well as in AT-2 cells in the parenchyma, consistent with previous studies[9-11]. Importantly, nasal epithelial cells, including two previously described clusters of goblet cells and one cluster of ciliated cells, show the highest expression among all investigated cells in the respiratory tree (**Fig. 1b**; left panel). We confirmed enriched *ACE2* expression in nasal epithelial cells in an independent scRNA-seq study that includes nasal brushings and biopsies. The results were consistent: we found the highest expression of *ACE2* in nasal secretory cells (equivalent to the two goblet cell clusters in the previous dataset) and ciliated cells (**Fig. 1b**; right panel).

In addition, scRNA-seq data from an *in vitro* epithelial regeneration system from nasal epithelial cells corroborated the expression of *ACE2* in goblet/secretory cells and ciliated cells in these air-liquid interface (ALI) cultures (**Extended Data Fig. 1**). Notably, the differentiating cells in ALI acquire progressively more *ACE2* (**Extended Data Fig. 1**). The results also suggest that this *in vitro* culture system may be biologically relevant for the study of SARS-CoV-2 pathogenesis.

It is worth noting that *TMPRSS2* was only expressed in a subset of *ACE2*$^{+}$ cells (**Extended Data Fig. 2**), suggesting that the virus might use alternative pathways. It was previously shown that SARS-CoV-2 could enter TMPRSS2^{-} cells using cathepsin B/L[7]. Indeed, other proteases were more promiscuously expressed than *TMPRSS2*, especially cathepsin B, which was expressed in more than 70%-90% of *ACE2*$^{+}$ cells

(**Extended Data Fig. 2**). However, while TMPRSS2 activity is documented to be important for viral transmission[13,14], the potential of cathepsin B/L or other proteases to functionally replace TMPRSS2 has not been determined.

We next asked whether the enriched expression of viral receptors and entry-associated molecules in the nasal region/upper airway might be relevant for viral transmissibility. Here, we assessed the expression of viral receptor genes that are used by other coronaviruses and influenza viruses in our datasets. We looked for *ANPEP* (used by HCoV-22944[15]) and *DPP4* (used by MERS-CoV45[16]), as well as the enzymes *ST6GAL1* and *ST3GAL4*, which are important for the synthesis of α(2,6)-linked and α(2,3)-linked sialic acids recognized by influenza viruses[17]. Notably, their expression distribution coincided with viral transmissibility patterns based on a comparison to the basic reproduction number (R_0), which estimates the number of people who can become infected from a single infected person. The skewed distribution of the receptors/enzymes towards the upper airway is observed in viruses with higher R_0/infectivity, including those of SARS-CoV/SARS-CoV-2 (R_0 ~ 1.4-5.0[18-20]), influenza (mean R_0 ~1.347[21]) and HCoV-229E (unidentified R_0; associated with common cold). This distribution is in distinct contrast with that of *DPP4*, the receptor for MERS-CoV (R_0 ~0.3-0.8[22]), a coronavirus with limited human-to-human transmission[23], in which expression skews towards lower airway/lung parenchyma (**Fig. 2a**). Therefore, our data highlight the possibility that viral transmissibility is dependent on the spatial distribution of receptor accessibility along the respiratory tract.

To gain more insight into the expression patterns of genes associated with *ACE2*, we performed Spearman's correlation analysis with Benjamini-Hochberg-adjusted *p*-values to identify genes associated with *ACE2* across all cells within the lung epithelial cell datasets. While the correlation coefficients are relatively low (< 0.12), likely due to low expression of *ACE2* and technical noise and dropout effects, the expression pattern of the top 50 *ACE2*-correlated genes across the respiratory tree is consistent with that of *ACE2*, with a skewed expression towards upper airway cells (**Fig. 2b** and **Extended Data Fig. 3a,b**). Interestingly, while some of the genes are associated with carbohydrate metabolism, possibly due to their role in goblet cell mucin synthesis, a number of genes associated with immune functions including innate and antiviral immune functions, are over-represented in the rank list, including *IDO1*, *IRAK3*, *NOS2*, *TNFSF10*, *OAS1*, and *MX1* (**Fig. 2b** and **Supplementary Table 1**). Expression of these genes is highest in nasal goblet 2 cells (**Fig. 2b**), consistent with the phenotype previously described. Nonetheless, nasal goblet 1 and nasal ciliated 2 cells also significantly express these genes (**Fig. 2b**). Given their environmental exposure and high expression of receptor/receptor-associated enzymes (**Fig. 2a**), it is plausible that nasal epithelial cells are conditioned to express these immune-associated genes to reduce viral susceptibility.

In this study, we explored multiple scRNA-seq datasets generated within the HCA consortium and other resources, and found that the SARS-CoV-2 entry receptor *ACE2* and viral entry-associated protease *TMPRSS2* are highly expressed in nasal goblet and ciliated cells. This finding implicates these cells as loci of original infection and possible reservoirs for dissemination within and between individuals. Co-expression in other barrier surface tissues could also suggest further investigation

into alternative transmission routes. For example, the co-expression in esophagus, ileum, and colon could explain viral fecal shedding observed clinically[24], with implications for potential fecal-oral transmission, whereas the co-expression in superficial conjunctival cells could explain an ocular phenotype observed in a small portion of COVID-19 patients[25] with the potential of spread through the nasolacrimal duct.

The results confirmed the expression of *ACE2* in multiple tissues shown in prior studies[10,11] with added information on tissues not previously investigated, including nasal epithelium and cornea, and its co-expression with *TMPRSS2*. We clearly detected nasal *ACE2* mRNA expression, for which protein confirmation is needed to resolve conflicting results in literature[8,12]. Our findings may have significant implications for understanding viral transmissibility, considering that the primary viral transmission is through infectious droplets. Moreover, as SARS-CoV-2 is an enveloped virus, its release does not require cell lysis. Thus, the virus might exploit existing secretory pathways in nasal goblet cells sustained at a pre-symptomatic stage. These discoveries could have translational implications. For example, given that nasal carriage is likely to be a key feature of transmission, drugs/vaccines administered intranasally could be highly effective in limiting spread.

This is the first collaborative effort by a Human Cell Atlas Biological Network (the Lung), and illustrates the opportunities from integrative analyses of Human Cell Atlas data, with future examples of consortium work expected soon.

Acknowledgements

We are grateful to Cori Bargmann, Jeremy Farrar, and Sarah Aldridge for stimulating discussions. We thank Jana Eliasova (scientific illustrator) for support with the design of figures, Sarah Sansum for support in document processing, and Martin Prete, Vlad Kiselev and the Wellcome Sanger Cellular Genetics IT Team, as well as Paul Bevan, for support with setting up the website portal. The human embryonic and fetal material was provided by the Joint MRC/Wellcome (MR/R006237/1) Human Developmental Biology Resource (www.hdbr.org).

This publication is part of the Human Cell Atlas - www.humancellatlas.org/publications.

This work was supported by the Wellcome Sanger Institute core funding (WT206194) and the Wellcome Strategic Scientific Support award “Pilot projects for the Human Cell Atlas” (WT211276/Z/18/Z), a Seed Network grant from the Chan Zuckerberg Initiative to P.B., T.D., T.E.D., O.E., P.H., N.H., N.K., M.K., K.B.M., A.M., M.C.N., M.N., D.P., J.R., P.R.T., S.Q., A.R., O.R., M.S., J.S., J.G.S., C.E.S., H.B.S., D.S., A.T., J.W. and K.Z. and by the European Union’s H2020 research and innovation program under grant agreement No 874656 (discovAIR) to P.B., A.B., M.K., S.L., J.L., K.B.M., M.C.N., K.S.P., C.S., H.B.S., J.S., F.J.T. and M.vd.B. W.S. acknowledges funding from the Newton Fund, Medical Research Council (MRC), The Thailand Research Fund (TRF), and Thailand’s National Science and Technology Development Agency (NSTDA). M.C.N acknowledges funding from GSK Ltd, Netherlands Lung Foundation project no. 5.1.14.020 and 4.1.18.226. T.D. acknowledges funding from HubMap consortium and Stanford Child Health Research Institute- Woods Family Faculty Scholarship. T.E.D. acknowledges funding from HubMap. P.H. acknowledges funding from LENDULET-BIOMAG Grant (2018-342) and the European Regional Development Funds (GINOP-

2.3.2-15-2016-00006, GINOP-2.3.2-15-2016-00026, GINOP-2.3.2-15-2016-00037). J.L.B. acknowledges funding from Medical Research Council (MRC), and the UK Regenerative Medicine Platform (MR/ 5005579/1). P.B. acknowledges funding from Fondation pour la Recherche Médicale (DEQ20180339158), Agence Nationale de la Recherche (UCAJEDI, ANR-15-IDEX-01; SAHARRA, ANR-19-CE14-0027; France Génomique, ANR-10-INBS-09-03), and Conseil Départemental des Alpes Maritimes (2016-294DGADSH-CV; 2019-390DGADSH-CV). N.E.B. and J.K. acknowledge funding from NIH grant R01HL145372 and DOD grant W81XWH1910416. I.G. acknowledges funding from NIH (5R24HD000836) and the Eunice Kennedy Shriver National Institute of Child Health and Human. N.H., J.G.S. and C.E.S. acknowledge funding by the Leducq foundation. N.H. is recipient of an ERC Advanced Grant. J.K. acknowledges funding from NIH grant K08HL130595 and the Doris Duke Charitable Foundation. N.K. acknowledges funding from NIH grants R01HL127349, U01HL145567 and an unrestricted grant from Three Lakes Foundation. M.K. acknowledges HHMI and Wall Center for Pulmonary Vascular Disease. H.L. acknowledges funding from National Research Foundation of Korea. K.M. acknowledges funding from Wellcome Trust. A.M. acknowledges funding from NIH grants HL135124, AG049665 and AI135964. M.Z.N. acknowledges funding from Rutherford Fund Fellowship allocated by the Medical Research Council and the UK Regenerative Medicine Platform (MR/ 5005579/1 to M.Z.N.). M.Z.N. and M.Y. have been funded by the Rosetrees Grant (Grant number M899). M.N. acknowledges funding from a BHF/DZHK grant and British Heart Foundation (PG/16/47/32156). J.O.-M. acknowledges funding from Richard and Susan Smith Family Foundation. D.P. acknowledges funding from Alan and Sandra Gerry Metastasis and Tumor Ecosystems Center. J.P. acknowledges funding from National Health and Medical

Research Council. P.R.T. acknowledges funding from R01HL146557 from NHLBI/NIH. E.L.R. acknowledges funding from MRC MR/P009581/1 and MR/SO35907/1. A.R. and O. R. acknowledge HHMI, the Klarman Cell Observatory, and the Manton Foundation. K.S.-P. acknowledges NIHR Cambridge Biomedical Research Centre. C.S. acknowledges Swedish research Council, Swedish Cancer Society, and CPI. H.B.S. acknowledges German Center for Lung Research and Helmholtz Association. J.S. acknowledges Boehringer Ingelheim, by the German Research Foundation (DFG; EXC2151/1, ImmunoSensation2 - the immune sensory system, project number 390873048), project numbers 329123747, 347286815) and by the HGF grant sparse2big. A.K.S. acknowledges the Beckman Young Investigator Program, a Sloan Fellowship in Chemistry, the NIH (5U24AI118672), and the Bill and Melinda Gates Foundation. F.J.T. acknowledges the German Center for Lung Research. M.vd.B. acknowledges from Ministry of Economic Affairs and Climate Policy by means of the PPP. K.B.W. is funded by the University College London-Birkbeck MRC Doctoral Training Programme. J.W. and Y.Y. acknowledge NIH, U01 HL148856 LungMap Phase II. R.X. acknowledges the NIH (DK043351). H.Z. is supported by the National Key R&D Program (no. 2019YFA0801703) and National Natural Science Foundation of China (no. 31871370)

Author Contributions

W.S., N.H., C.B., and M.B. performed data analyses, initiated by S.A.T. and supervised by S.A.T. and M.C.N. R.Q., M.L., C.T.-L., H.M., D.R., F.S., M.Y., K.B.W., and the HCA Lung Biological Network generated and provided data. W.S, N.H. and the HCA Lung Biological Network interpreted the data. W.S., with significant input from

the HCA Lung Biological Network, especially M.C.N., M.vd.B., S.A.T., and K.B.M., wrote the paper. K.B.W., M.Y., and J.L.B. performed experiments for the revision process. All authors read the manuscript, offered feedback, and approved it before submission.

Competing Interests

N.K. was a consultant to Biogen Idec, Boehringer Ingelheim, Third Rock, Pliant, Samumed, NuMedii, Indaloo, Theravance, LifeMax, Three Lake Partners, Optikira in the last three years and received non-financial support from MiRagen. J.L. is a scientific consultant for 10X Genomics Inc. J.K. reports advisory board fees from Boehringer Ingelheim, nonfinancial study support from Genentech, and grant funding from Boehringer Ingelheim. A.R. is a co-founder and equity holder of Celsius Therapeutics, an equity holder in Immunitas, and an SAB member of ThermoFisher Scientific, Syros Pharmaceuticals, Asimov, and Neogene Therapeutics. O.R. is a co-inventor on patent applications filed by the Broad Institute to inventions relating to single cell genomics applications, such as in PCT/US2018/060860 and US Provisional Application No. 62/745,259. A.K.S. reports compensation for consulting and/or SAB membership from Merck, Honeycomb Biotechnologies, Cellarity, Cogen Therapeutics, Orche Bio, and Dahlia Biosciences. F.J.T. reports receiving consulting fees from Roche Diagnostics GmbH, and ownership interest in Cellarity Inc. S.A.T. was a consultant at Genentech, Biogen and Roche in the last three years. L.V. is a founder of Definigen and Bilitech, two biotech companies using hPSCs and organoid for disease modelling and cell-based therapy. F.S. is a founder of Bilitech, a biotechnology company using organoids for cell-based therapy.

Reference

1. World Health Organization. (2020). Retrieved from https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it
2. Chen, N.*, et al. Lancet* **395**, 507-513 (2020).
3. Zhu, N.*, et al*. *N Engl J Med* **382**, 727-733 (2020).
4. Zhou, P.*, et al. Nature*, **579**, 270-273 (2020).
5. Li, W.*, et al. Nature* **426**, 450-454 (2003).
6. Matsuyama, S.*, et al. Journal of virology* **84**, 12658-12664 (2010).
7. Hoffmann, M.*, et al. Cell* (2020).
8. Bertram, S.*, et al. PloS one* **7**, e35876 (2012).
9. Zhao, Y.*, et al. bioRxiv*, 2020.2001.2026.919985 (2020).
10. Zou, X.*, et al. Frontiers of medicine* (2020).
11. Qi, F., *et al. bioRxiv*, 2020.2002.2016.951913 (2020).
12. Hamming, I.*, et al.The Journal of pathology* **203**, 631-637 (2004).
13. Zhou, Y.*, et al. Antiviral research* **116**, 76-84 (2015).
14. Iwata-Yoshikawa, N.*, et al. Journal of virology* **93**(2019).
15. Yeager, C.L.*, et al. Nature* **357**, 420-422 (1992).
16. Raj, V.S.*, et al. Nature* **495**, 251-254 (2013).
17. Broszeit, F.*, et al. Cell Rep* **27**, 3284-3294.e3286 (2019).
18. Li, Q.*, et al. N Engl J Med* (2020).
19. Wallinga, J. & Teunis, P. *American journal of epidemiology* **160**, 509-516 (2004).
20. Riou, J. & Althaus, C.L. *Euro surveillance: bulletin Europeen sur les maladies transmissibles* **25**(2020).

Figure Legends

Fig. 1| Expression of *ACE2* and *TMPRSS2* across different tissues and its enrichment in nasal epithelial cells. **a**, RNA expression of SARS-CoV-2 entry receptor *ACE2* (first column), entry-associated protease *TMPRSS2* (second column), and their co-expression (third column) from multiple scRNA-seq datasets across different tissues. Raw expression values were normalized, log transformed and summarized by published cell clustering where available, or reproduced clustering annotated using marker genes and cell type nomenclature from the respective studies. The size of the dots indicates the proportion of cells in the respective cell type having greater-than-zero expression of *ACE2* (first column), *TMPRSS2* (second column) or

both (third column), while the colour indicates the mean expression of *ACE2* (first and third columns) or *TMPRSS2* (second column). **b**, Schematic illustration depicts the major anatomical regions in the human respiratory tree demonstrated in this study: nasal, lower airway, and lung parenchyma (left panel). Expression of *ACE2* is from airway epithelial cell datasets: Vieira Braga, Kar *et al.* 2019 (middle panel) and Deprez *et al.* 2019 (right panel). The datasets were retrieved from existing sources, and the cell clustering and nomenclature were retained based on the respective studies. For gene expression results in the dot plots: the dot size represents the proportion of cells within the respective cell type expressing the gene and the dot color represents the average gene expression level within the particular cell type.

Fig. 2| Respiratory expression of viral receptor/entry-associated genes and implications for viral transmissibility and genes associated with *ACE2* expression. a, Expression of *ACE2* (an entry receptor for SARS-CoV and SARS-CoV-2), *ANPEP* (an entry receptor for HCoV-229E), *ST6GAL1*/*ST3GAL4* (enzymes important for synthesis of influenza entry receptors), and *DPP4* (an entry receptor for MERS-CoV) from the airway epithelial datasets: Vieira Braga, Kar *et al*. 2019 (left panel) and Deprez *et al*. 2019 (right panel). The basic reproductive number (R0) for respective viruses, if available, are shown. **b**, Respiratory epithelial expression of the top 50 genes correlated with *ACE2* expression based on Spearman's correlation analysis (with Benjamini-Hochberg-adjusted *p*-values) performed on all cells within the Vieira Braga, Kar *et al.* airway epithelial dataset. The colored gene names represent genes that are immune-associated (GO:0002376: immune system process or GO:0002526: acute inflammatory response). For gene expression results in the dot plots: the dot size represents the proportion of cells within the respective cell type

expressing the gene and the color represents the average gene expression level within the particular cell type.

Methods

The datasets were retrieved from published and unpublished datasets in multiple human tissues, including airways[26,27], cornea (personal communication; Lako lab, Newcastle), skeletal muscle (personal communication, Teichmann lab, Wellcome Sanger Insitute and Zhang lab, Sun-Yat-Sen University, Guangzhou, China), ileum[28], colon[29], pancreas[30], liver[31], gallbladder (personal communication; Vallier lab, University of Cambridge), heart (Teichmann lab, Hubner lab/Berlin, Seidmanns/Harvard, and Noseda lab/Imperial College London), kidney[32], placenta/decidua[33], testis[34], prostate gland[35], brain[36], skin[37], retina[38], spleen[39], esophagus[39], and fetal tissues[40,41]. Raw expression values were normalized and log transformed. We retained the cell clustering based on the original studies when available.

For each dataset where per-cell annotation is not available, we re-processed the data from raw or normalized (whichever was deposited alongside the original publication) quantification matrix. The standard scanpy (version 1.4.3) clustering procedure was followed. When batch information is available, harmony package was used to correct batch effects in the PC space and the corrected PCs were used for computing nearest neighbour graphs. To re-annotate the cells, multiple clusterings of different resolutions were generated among which the one best matching the published clustering was picked and manual annotation was undertaken using marker genes described in the original publication. Full details can be found in analysis notebooks available at

github.com/Teichlab/covid19_MS1.

Illustration of the results was generated using scanpy and Seurat (version 3.1). For correlation analysis with *ACE2*, we performed the Spearman's correlation with statistical tests using the R Hmisc package (version 4.3-1) and the p values were adjusted with Benjamini-Hochberg method with the R stats package (version 3.6.1) on the Vieira Braga, Kar *et al*. airway epithelial dataset and the Deprez *et al*. airway dataset. We also tested multiple additional approaches, including Kendall's correlation, data transformation by sctransform function in the Seurat package, and data imputation by the Markov Affinity-based Graph Imputation of Cells (MAGIC) algorithm, to compare correlation results. While the imputation significantly improved the correlations, the top genes correlated with *ACE2* are largely the same as the analysis done on un-imputed data. With the uncertainty of the extent imputation artificially distorted the data, we reported the results with no imputation even though the correlations are low. The correlation coefficients for all genes are included as **Supplementary Data 1**. The top 50 genes in each dataset were characterized based on Gene Ontology classes from the Gene Ontology (GO) database and associated pathways in PathCards from the Pathway Unification database.

Data Availability Statement

The published datasets can be found as followed: pulmonary airways (European Genome-phenome Archive: EGAS00001001755, EGAS00001002649; lungcellatlas.org and www.genomique.eu/cellbrowser/HCA), ileum (NCBI: GSE134809), colon (Single Cell Portal: SCP259; portals.broadinstitute.org/single_cell), pancreas (NCBI: GSE84133), liver (NCBI:

GSE115469), kidney (www.kidneycellatlas.org), placenta/decidua (EBI Array Express: E-MTAB-6701; maternal-fetal-interface.cellgeni.sanger.ac.uk), testis (NCBI: GSE120508), brain (www.gtexportal.org/home/data-sets), retina (NCBI: GSE135922), skin (European Genome-phenome Archive: EGAS00001002927), spleen and esophagus (tissuestabilitycellatlas.org) and fetal tissues (Array Express: E-MTAB-7407 and E-MTAB-8581; developmentalcellatlas.ncl.ac.uk).

All of the published datasets and relevant data from unpublished sources in this study can be visualized and assessed through a website portal (covid19cellatlas.org).

Code Availability Statement

Analysis notebooks are available at github.com/Teichlab/covid19_MS1.

Methods-only References

21. Coburn, B.J., Wagner, B.G. & Blower, S. *BMC medicine* **7**, 30 (2009).
22. Kucharski, A.J. & Althaus, C.L. *Euro surveillance : bulletin Europeen sur les maladies transmissibles* **20**, 14-18 (2015).
23. Killerby, M.E., *et al. Emerging infectious diseases* **26**, 191-198 (2020).
24. Xu, Y.*, et al. Nature medicine* (2020).
25. Guan, W.J.*, et al. N Engl J Med* (2020).
26. Vieira Braga, F.A.*, et al. Nature medicine* **25**, 1153-1163 (2019).
27. Deprez, M.*, et al. bioRxiv*, 2019.2012.2021.884759 (2019).
28. Martin, J.C.*, et al. Cell* **178**, 1493-1508.e1420 (2019).
29. Smillie, C.S.*, et al. Cell* **178**, 714-730.e722 (2019).
30. Baron, M.*, et al.. Cell systems* **3**, 346-360.e344 (2016).
31. MacParland, S.A.*, et al. Nature communications* **9**, 4383 (2018).
32. Stewart, B.J.*, et al. Science* **365**, 1461-1466 (2019).
33. Vento-Tormo, R.*, et al. Nature* **563**, 347-353 (2018).
34. Guo, J.*, et al. Cell research* **28**, 1141-1157 (2018).
35. Henry, G.H.*, et al. Cell Rep* **25**, 3530-3542.e3535 (2018).
36. Habib, N.*, et al. Nature methods* **14**, 955-958 (2017).
37. Cheng, J.B.*, et al. Cell Rep* **25**, 871-883 (2018).
38. Voigt, A.P.*, et al. Experimental eye research* **184**, 234-242 (2019).
39. Madissoon, E.*, et al. Genome Biol* **21**, 1 (2019).
40. Popescu, D.M.*, et al. Nature* **574**, 365-371 (2019).
41. Park, J.E.*, et al. Science* **367**(2020).

The HCA Lung Biological Network

*The HCA Lung Biological Network are: Nicholas E. Banovich[1], Pascal Barbry[2], Alvis Brazma[3], Joseph Collin[4], Tushar J. Desai[5], Thu Elizabeth Duong[6,7], Oliver Eickelberg[8], Christine Falk[9,10], Michael Farzan[11], Ian Glass[12], Ravindra K. Gupta[13], Muzlifah Haniffa[4,14,15], Peter Horvath[16,17], Norbert Hubner[18], Deborah Hung[20,21,22], Naftali Kaminski[23], Mark Krasnow[24], Jonathan A. Kropski[25], Malte Kuhnemund[26], Majlinda Lako[4], Haeock Lee[27], Sylvie Leroy[2,28], Sten Linnarson[29], Joakim Lundeberg[30], Kerstin B. Meyer[14], Chichau Miao[14], Alexander J. Misharin[31], Martijn C. Nawijn[32], Marko Z. Nikolic[33], Michela Noseda[34,35], Jose OrdovasMontanes[36,37,38,39], Gavin Oudit[40], Dana Pe'er[41], Joseph Powell[42,43], Steve Quake[44], Jay Rajagopal[39,45], Purushothama Rao Tata[46], Emma L. Rawlins[47], Aviv Regev[38,48,49,50], Paul A. Reyfman[51], Orit Rozenblatt-Rosen[38,48], Kourosh Saeb-Parsy[52], Christos Samakovlis[30,53], Herbert B. Schiller[54], Joachim L. Schultze[55,56], Max Seibold[57], Christine E. Seidman[20,49,58], Jonathan G. Seidman[20], Alex K. Shalek[38,59,60], Douglas Shepherd[61], Jason Spence[62], Avi Spira[63,64], Xin Sun[65], Sarah A. Teichmann[14,66], Fabian J. Theis[67], Alexander Tsankov[68], Ludovic Vallier[69], Maarten van den Berge[70], Jeffrey Whitsett[71], Ramnik Xavier[21,38], Yan Xu[72], Laure-Emmanuelle Zaragosi[2], Darin Zerti[4,73], Hongbo Zhang[74], and Kun Zhang[75].

[1]Translational Genomics Research Institute, Phoenix, AZ

[2]Université Côte d'Azur, CNRS, IPMC, Sophia-Antipolis 06560, France

[3]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK

[4]Biosciences Institute, Faculty of Medical Sciences, Newcastle University, International Centre for Life, Newcastle upon Tyne, UK

[5]Department of Medicine and Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA

[6]Department of Pediatrics Division of Respiratory Medicine, University of California San Diego [7]Rady Children's Hospital San Diego, San Diego, CA 92123, USA

[8]Division of Pulmonary Sciences and Critical Care Medicine, Department of Medicine, University of Colorado, Anschutz Medical Campus, Aurora, CO, US

[9]Institute of Transplant Immunology, Hannover Medical School, MHH, Carl-Neuberg Str. 1, 30625 Hannover, Germany,

[10]German Center for Infectious Diseases DZIF, TTU-IICH 07.801

[11]Department of Immunology and Microbiology, The Scripps Research Institute, Jupiter, Florida, USA (33458)

[12]Department of Pediatrics, Genetic Medicine, University of Washington, Seattle, Washington

[13]Cambridge Institute for Therapeutic Immunology and Infectious Diseases, Jeffrey Cheah Biomedical Centre, University of Cambridge, Cambridge CB2 0AW, UK

[14]Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, UK

[15]Department of Dermatology and NIHR Newcastle Biomedical Research Centre, Newcastle Hospitals NHS Foundation Trust, Newcastle upon Tyne NE2 4LP, UK

[16]Synthetic and Systems Biology Unit, Biological Research Centre (BRC), Szeged, Hungary

[17]Institute for Molecular Medicine Finland, University of Helsinki

[18]Cardiovascular and Metabolic Sciences, Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), 13125 Berlin, Germany

[19]DZHK (German Centre for Cardiovascular Research), Partner Site Berlin, 13347 Berlin, Germany

[20]Department of Genetics,Harvard Medical School, Boston, MA 02115, USA

[21]Department of Molecular Biology at Massachusetts General Hospital, Boston, MA 02114, USA

[21]Infectious Disease and Microbiome Program, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

[23]Pulmonary, Critical Care and Sleep Medicine, Yale University School of Medicine, New Haven, CT 06520, USA

[24]Department of Biochemistry and Wall Center for Pulmonary Vascular Disease, Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA 94305, USA

[25]Division of Allergy, Pulmonary and Critical Care Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, Department of Veterans Affairs Medical Center, Nashville, TN; Department of Cell and Developmental Biology, Vanderbilt University, Nashville, TN

[26]Cartana AB, Nobels vag 16, 17165 Stockholm, Sweden

[27]Department of Biomedicine and Health Sciences, The Catholic University of Korea, Seoul, Korea

[28]Institut de Pharmacologie Moléculaire et Cellulaire, Sophia Antipolis, France

[29]Division of Molecular Neurobiology, Department of Medical Biochemistry and Biophysics, Karolinska Institute

[30]SciLifeLab, Stockholm University, 17121, Solna, Sweden

[31]Division of Pulmonary and Critical Care Medicine, Northwestern University, Chicago, Illinois 60611, USA

[32]Department of Pathology and Medical Biology, University of Groningen, GRIAC Research Institute, University Medical Center Groningen, 9713 AV Groningen, Netherlands

[33]UCL Respiratory, Division of Medicine, University College London, WC1E 6JF, London, UK

[34]National Heart and Lung Institute, Imperial College London, UK

[35]British Heart Foundation Centre for Research Excellence and Centre for Regenerative Medicine, Imperial College London, UK

[36]Division of Gastroenterology Boston Children's Hospital, Boston, MA 02115, USA

[37]Program in Immunology, Harvard Medical School, Boston, MA 02115, USA

[38]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

[39]Harvard Stem Cell Institute, Cambridge, MA 02138, USA

[40]Canada Research Chair in Heart Failure, Division of Cardiology, 2C2 Walter Mackenzie Health Sciences Centre, Edmonton, Alberta, T6G 2B7, Canada

[41]Computational and Systems Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA

[42]Garvan-Weizmann Centre for Cellular Genomics, Garvan Institute of Medical Research, Sydney, NSW, Australia

[43]UNSW Cellular Genomics Futures Institute, University of New South Wales, Sydney, NSW, Australia

[44]Depts of Bioengineering and Applied Physics, Stanford University, and the Chan Zuckerberg Biohub, Stanford University, Stanford, CA 94305, USA

[45]Center for Regenerative Medicine, Massachusetts General Hospital, Boston, MA 02114, Boston

[46]Department of Cell Biology, Regeneration Next Initiative, Duke University School of Medicine, Durham, NC 27710, USA

[47]Wellcome Trust/CRUK Gurdon Institute and Department Physiology, Development and Neuroscience, University of Cambridge, Cambridge, CB2 1QN, UK

[48]Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

[49]Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA

[50]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142, USA

[51]Northwestern University Feinberg School of Medicine, Division of Pulmonary and Critical Care Medicine

[52]Department of Surgery, University of Cambridge and NIHR Cambridge Biomedical Research Centre, CB2 0QQ, UK

[53]Cardiopulmonary Institute, Justus Liebig University; Giessen Germany

[54]Comprehensive Pneumology Center (CPC) / Institute of Lung Biology and Disease (ILBD), Helmholtz Zentrum München, Member of the German Center for Lung Research (DZL), Munich, Germany

[55]Department for Genomics & Immunoregulation, LIMES-Institute, University of Bonn, 53115 Bonn, Germany Department of Pediatrics; Center for Genes, Environment, and Health

[56]PRECISE Platform for Single Cell Genomics & Epigenomics, German Center for Neurodegenerative Diseases and University of Bonn, Bonn, Germany

[57]Department of Pediatrics, Center for Genes, Environment, and Health, National Jewish Health; Denver, CO 80206

[58]Cardiovascular Division, Brigham & Women's Hospital, Boston, MA 02115, USA

[59]Ragon Institute of MGH, MIT, and Harvard, Cambridge, MA, USA

[60]Institute for Medical Engineering and Science (IMES), Koch Institute for Integrative Cancer Research, and Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA, USA;

[61]Center for Biological Physics and Department of Physics, Arizona State University, Tempe, AZ 85287, USA

[62]Department of Internal Medicine, Gastroenterology, University of Michigan Medical School, Ann Arbor, MI 48109, USA; Department of Cell and Developmental Biology, University of Michigan Medical School, Ann Arbor, MI 48109, USA; Department of Biomedical Engineering, University of Michigan College of Engineering, Ann Arbor, MI 48109, USA.

[63]Department of Medicine, Boston University School of Medicine, Boston, MA, USA

[64]Johnson & Johnson Innovation, Cambridge, MA, USA.

[65]Department of Pediatrics, Department of Biological Sciences, University of California SD, 9500 Gilman Dr. MC0766, San Diego, CA 92093, USA

[66]Theory of Condensed Matter Group, Cavendish Laboratory/Department of Physics, University of Cambridge, Cambridge CB3 0HE, UK

[67]Institute of Computational Biology, Helmholtz Zentrum München and Departments of Mathematics and Life Sciences, Technical University Munich, Germany

[68]Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

[69]Wellcome and MRC Cambridge Stem Cell Institute, Jeffrey Cheah Biomedical Centre, Biomedical Campus, Puddicombe Way, Cambridge CB2 0AW, UK; Department of Surgery, Cambridge Biomedical Campus, Hills Rd, Cambridge, CB2 0QQ, UK

[70]Department of Pulmonary diseases and tuberculosis, University of Groningen, GRIAC Research Institute, University Medical Center Groningen, 9713 AV Groningen, Netherlands

[71]Cincinnati Children's Hospital Medical Center, Cincinnati, OHIO

[72]Divisions of Pulmonary Biology and Biomedical Informatics; Perinatal Institute, Cincinnati Children's Hospital Medical Center; University of Cincinnati College of Medicine

[73]Microscopy Centre and Department of Applied Clinical Sciences and Biotechnology, University of L'Aquila, via Vetoio, 67100 Coppito, L'Aquila, Italy

[74]Key Laboratory for Stem Cells and Tissue Engineering, Ministry of Education and Department of Histology and Embryology of Zhongshan School of Medicine, Sun Yat-Sen University, Guangzhou 510080, China

[75]UCSD Department of Bioengineering, 9500 Gilman Drive, MC0412, PFBH402, La Jolla, CA 92093, USA

Pascal Barbry, Alexander Misharin, Martijn Nawijn and Jay Rajagopal serve as the coordinators for the HCA Lung Biological Network.